(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 November 2001 (08.11.2001)

PCT

(10) International Publication Number
WO 01/84875 A2

(51) International Patent Classification[7]: H04Q 11/00

(21) International Application Number: PCT/US01/13508

(22) International Filing Date: 27 April 2001 (27.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/563,338 3 May 2000 (03.05.2000) US

(71) Applicant: **NOKIA, INC.** [US/US]; 6000 Connection Drive, Irving, TX 75039 (US).

(72) Inventors: **CHASKAR, Hemant**; 111 Lucust Street, Apartment 40-C-1, Woburn, MA 01801 (US). **VERMA, Sanjeev**; 2 Kimball Court, Apartment 213, Woburn, MA 01801 (US). **RAVIKANTH, R.**; 1504 Stearns Hill Road, Waltham, MA 02451 (US). **DIXIT, Sudhir, S.**; 12 Westerly Road, Weston, MA 02493 (US).

(74) Agents: **BRUNDIDGE, Carl, I.** et al.; Antonelli, Terry, Stout & Kraus, LLP, Suite 1800, 1300 N. Seventeenth Street, Arlington, VA 22209 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/84875 A2

(54) Title: ROBUST TRANSPORT OF IP TRAFFIC OVER WDM USING OPTICAL BURST SWITCHING

(57) Abstract: A technique for selecting the offset between data bursts and their respective control packets in an optical burst switching arrangement includes: randomly generating a plurality of tokens; receiving a plurality of sequentially generated data bursts; and receiving a plurality of control packets, each control packet corresponding to a respective one of the plurality of data bursts. One of the plurality of control packets is first outputted and its corresponding respective data burst is then output at a time corresponding to the generation of the first of the plurality of tokens which occurs at a time in which no data burst is being outputted, the outputting of the data burst being offset from the output of its corresponding respective control packet by a time period. The average rate at which the plurality of data bursts are outputted may be equal to the reciprocal of the mean of the probability distribution used to generate the plurality of tokens. The plurality of tokens may be randomly generated according to a Poisson process.

ROBUST TRANSPORT OF IP TRAFFIC OVER WDM USING OPTICAL BURST SWITCHING

TECHNICAL FIELD

BACKGROUND ART

The present invention relates to the routing of Internet protocol traffic using optical burst switching and more particularly the present invention relates to selecting the offset between a control packet and a data burst to optimize the traffic flow.

Rapid growth in the volume of Internet traffic over the last decade has generated a considerable amount of interest in devising new high-speed transmission and switching technologies. Wavelength division multiplexing can support a number of simultaneous high-speed channels on a single optical fiber and can thereby provide an enormous bandwidth at the physical layer. In order to exploit this bandwidth to meet the future traffic requirements, higher layer communication protocols must be developed to make efficient use of the transmission capacity of such optical fiber arrangements.

Presently, wavelength division multiplexing deployment is mostly point-to-point and the Internet protocol routers and asynchronous transfer mode switches in such a backbone still use electronic processing, such as header processing for routing including table lookups or label swapping and data-processing for multiplexing including mapping Internet protocol packets into asynchronous transfer mode cells before transporting over wavelength division multiplexed links using SONET frames. Since the operating speed of electronic devices is considerably slower than the transmission speed provided by the optical wavelength division multiplexing layer, the optical-electronics-optical conversion at the intermediate nodes in the data path of the wavelength division multiplexing layer should be eliminated.

Ideally, at all-optical packet switch arrangement so as to eliminate the electronics entirely would eliminate the speed bottleneck. However, the present limitations of optical buffering, among other things, preclude entirely optical packet switching today.

As a presently implementable alternative to all-optical packet switches, optical burst switching still allows the switching of data channels in the optical domain by performing

resource allocation in the electronic domain. In optical burst switching, a control packet precedes every data burst and the control packet and the corresponding data burst are launched at the source at points in time separated by an offset. The offset is determined at the time that the control packet is launched at the source. The control packet includes information required to route the data burst through the network and also includes the length of the corresponding data burst and its offset value. The control packet is processed electronically at each of the intermediate nodes for making routing decisions and the switching fabric at each node is configured accordingly to switch to the data burst that is expected to arrive after a time interval corresponding to the offset field of the control packet. Thus, the data burst is entirely optically switched to eliminate the electronic bottleneck.

DISCLOSURE OF THE INVENTION

The present invention determines the offset between the control packet and the data burst. While the offset must be at least as large as the sum of the processing times for the control packet at each of the intermediate nodes, the present invention determines an offset which can reduce the contention among reservations requested by different control packets for different flows so as to significantly improve the performance of the optical burst switching arrangement. That is, the present invention determines the offsets of successive data bursts from their control packets which results and reliable operation of the optical burst switching wavelength division multiplexing network. In the present invention, the data bursts are released into the network at random times to improve the burst blocking performance at individual nodes in the networkwide-contention-prone optical burst switching wavelength division multiplexing backbone.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and a better understanding of the present invention will become apparent from the following detailed description of example embodiments and the claims when read in connection with the accompanying drawings, all forming a part of the disclosure of this invention. While the foregoing and following written and illustrated disclosure focuses on disclosing example embodiments of the invention, it should be clearly understood that the same is by way of illustration and example only and the invention is not limited thereto. The spirit and scope of the present invention are limited only by the terms of the appended claims.

The following represents brief descriptions of the drawings, wherein:

Figure 1 is a block diagram of optical burst switching network.

Figure 2 is a block diagram of an optical network transmitting Internet protocol data over a wavelength division multiplexing transmission backbone using multiprotocol label switching.

Figure 3 is a block diagram illustrating a shaping interface between the Internet protocol layer and the wavelength division multiplexing optical layer.

Figure 4 is a timing diagram of the data bursts before and after shaping and the control packets.

Figure 5 illustrates the burst blocking probability as a function of the offered load for an interface supporting various numbers of wavelengths.

Figure 6 illustrates the mapping of Internet protocol classes onto optical label switched paths.

Figure 7 illustrates quality of service differentiation in terms of burst blocking probability.

BEST MODE FOR CARRYING OUT THE INVENTION

Before beginning a detailed description of the subject invention, mention of the following is in order. When appropriate, like reference numerals and characters may be used to designate identical, corresponding or similar components in differing drawing figures. Furthermore, in the detailed description to follow, example sizes/models/values/ranges may be given, although the present invention is not limited thereto. Still furthermore, the clock and timing signal figures are not drawn to scale, and instead, exemplary and critical time values are mentioned when appropriate. In addition, well-known power connections to the components have not been shown within the drawing figures for simplicity of illustration and discussion and so as not to obscure the invention.

Figure 1 is a block diagram of an optical burst switching network. Interfaces 150, 151,152, and 153 are disposed along the optical switching path. A data burst 160 is first launched at the source and is followed by the control packet 170 which is launched after an offset 180 from the start of the data burst 160. In the control unit 100, of interface i, the control packet 170 is first converted from an optical signal into an electrical signal by an optical-electrical converter 110. The output of the optical-electrical converter 110 controls label swapping 120 and a wavelength scheduler 130 and the resultant processed output is converted from electrical signal into an optical signal in an electrical-optical converter 140 whose output is used to switch the data burst 160 that is expected to arrive after the time interval contained within the offset field of the control packet 170.

Figure 2 is a block diagram of an optical network transmitting Internet protocol data over a wavelength division multiplexing transmission backbone using multiprotocol label switching.

Internet protocol routers 201,202, and 203 are disposed at the ingress of the network and Internet protocol routers 230 and 240 are disposed at the egress of the network. Intermediate nodes 210, 211, and 212 are disposed between the ingress routers and the egress routers. Data bursts are assembled at the ingress routers and delivered to the ingress routers via the intermediate nodes in an all-optical domain. It is assumed that there is no buffering of data bursts at the intermediate nodes. Semi-permanent data pipes can be set up between different ingress-egress pairs using multiprotocol label switching.

An Internet protocol routing and sorting engine causes a major bottleneck at high transmission speeds due to its processing requirements. Multiprotocol label switching is a forwarding technique which uses the labels associated with packets to make packet forwarding decisions at the network nodes rather than by conventional destination-based hop-by-hop forwarding arrangements. In multiprotocol label switching, the space of all possible fowarding options is partitioned into forwarding equivalence classes. For example, all of the packets destined for a given egress node which have the same quality of service requirement may belong to the same forwarding equivalence class. The packets are labeled at the ingress in accordance with the forwarding equivalence class with which they belong. Each of the intermediate nodes uses the label of an incoming packet to determine its next hop and also performs label swapping, that is, it replaces the incoming label with the new outgoing label which identifies the respective forwarding equivalence class for the downstream node. Such a label-based forwarding technique reduces the processing overhead required for routing at the intermediate nodes, thereby improving their packet forwarding performance and scalability. Furthermore, the label swapping processed used by multiprotocol label switching creates a multipoint-to-point routing tree rather than a routing mesh used in conventional networks. Multiprotocol label switching also provides constraint-based routing in which the ingress node can establish an explicit route through the network rather than inefficiently carrying the explicit route in each packet. Instead, multiprotocol label switching allows the explicit route to be carried only at the time that the label switched path is set up. The subsequent packets traversing this path are forwarded using packet labels.

The control packets that precede the data bursts can be used to carry multiprotocol label switching labels. The intermediate nodes use the labels in incoming control packets to set up the switch fabric, thereby allowing the data burst to be forwarding in an all-optical domain.

The data bursts are freed from carrying any labels. Multiprotocol label switching thereby allows the establishment of data pipes through an optical burst switching wavelength division multiplexing domain between different ingress-egress pairs. These data pipes will be referred to hereinafter as optical LSP's (label switched paths).

Figure 3 is a block diagram illustrating a shaping interface between the Internet protocol layer and the wavelength division multiplexing optical layer. A medium access control interface 350 is shown as being disposed between the Internet protocol layer 300 and the optical layer 370. The medium access control layer 350, which includes a burst assembly 313 and a burst scheduler 324 and shaper 335, assembles data bursts from the packets generated by the Internet protocol layer 300 for each optical LSP. Various quality of service considerations, such as DiffServ can be used at the Internet protocol layer 300 before the packets are released to the medium access control layer 350.

The optical LSP 340 of the medium access control 350 appears to the Internet protocol layer 300 as a reliable data pipe that guarantees certain burst blocking probability to the Internet protocol layer 300. The delay across the optical LSP 340 has to components, namely, a fixed component which mainly consists of the propagation delay across the optical backbone and a variable component which depends on the design of the shaper 335 and burst assembler 313 at the ingress router.

An output interface of a node in the optical burst switching wavelength division multiplexing domain received traffic from different LSP's. This creates the potential for contention among different LSP's. When the control packets from different LSP's request reservations for their data bursts on a particular wavelength of a given output interface for time intervals which overlap each other, hereinafter referred to as burst blocking, only one of these requests can be granted. Accordingly, some control packets must be dropped at that interface and is in turn results in the loss of data bursts corresponding to the dropped control packets.

The data loss rate due to burst blocking will be large if the reservation requests arriving at a given output interface from different optical LSP's are time correlated. Furthermore, due to the unpredictability of traffic, it is difficult to always guarantee a low burst blocking probability.

In accordance with a present invention, a shaping mechanism, including the shaper 335, for example, is incorporated in the medium access control layer 350 at the ingress thereof which facilitates the determination of offsets for the successive data bursts so as to maintain a low burst blocking probability at all times in the optical burst switching wavelength division multiplexing layer. The shaping mechanism removes the time correlation among the

reservation requests of different optical LSP's and enforces predetermined statistics on the data stream entering the optical burst switching wavelength division multiplex layer, irrespective of the statistics of the Internet protocol layer that resides over it. Due to the bufferless operation of the optical burst switching wavelength division multiplexing layer, this statistic is invariant even if the burst stream traverses multiple nodes in the optical burst switching wavelength division multiplexing domain. This results in a robust layer in which the probability of burst blocking is always maintained below a low predetermined value irrespective of the operation of the sources generating the data packets.

As shown in figure 3, the stream of data bursts to be sent over a particular optical LSP is inputted to the shaper 335 which determines the value of the offset between the data burst and its corresponding control packet. The shaper 335 then forms the control packet and transmits it over the control channel. The control packet includes information such as the offset value between the data burst and the corresponding control packet, the length of the data burst and the routing label. The shaper 335 causes the data burst to be outputted to the optical burst switching wavelength division multiplexing layer after the control packet has been transmitted, the data burst being outputted a period of time equal to the offset value after the control packet has been transmitted.

The offset between a data burst and its corresponding control packet has two parts, namely, a constant part to account for the processing time of the control packet at the intermediate nodes and a variable part which is varied from burst to burst to lower the probability of burst blocking. The variable part of the offset for the ith data burst, denoted by δ_i, is determined as follows:

Let $T_0=0$, T_1, T_2, ...denote the times of occurrences of points of a random point process in which the time periods between the occurrences of successive points (i.e., T_i-T_{i-1}, for $i \geq 1$) are independently and identically distributed according to the probability distribution $F(.)$. Let $T_0(\omega)=0, T_1(\omega)$, $T_2(\omega)$, ...denotes a particular realization (sample path) of this random point process. If the ith data burst arrives at the shaper at time a_i, and that the $(i-1)$th data burst is released at $T_{k_{i-1}}(\omega)$, then the ith burst is released at time $T_{k_i}(\omega)$,

where $T_{k_i}(\omega)$ is the first point after $T_{k_{i-1}}(\omega)$ satisfying the following inequality:

$$T_{k_{i-1}}(\omega) - T_{k_{i-1}}(\omega) \geq L_{i-1} \text{ and } T_{k_i}(\omega) \geq a_i.$$

Hence, the offset between the ith data burst and the control packet corresponding to it is determined by the following equation:

$$\delta_{ii} = T_{k_i}(\omega) - a_i.$$

The shaping scheme described above (see Figure 4) is equivalent to leaky bucket regulator with *no* buffering provided for tokens, and in which tokens arrive at $T_0(\omega)$,$T_1(\omega)$, $T_2(\omega)$, ...

Figure 4 is a timing diagram of the data bursts before and after shaping and the control packets. Figure 4(A) illustrates the data bursts L1, L2, L3, and L4 before shaping by the shaper 335. Figure 4(B) illustrates the arrival of the tokens T1, T2, T3, T4 and T5. Figure 4(C) illustrates the data bursts d1,d2, d3, and d4. Figure 4(D) illustrates the control packets corresponding to the data bursts of Figure 4(A).

As illustrated in Figure 4, the data bursts are of random length and occur at seemingly random times which may in fact be time correlated. As noted above, the token arrivals have been chosen so as to be randomly distributed. The data burst L1 is shaped by the shaper 335 into the shaped data burst d1 by having its transmission beginning at the arrival of token T1 which occurs at a time $\delta 1$ after the start of the control packet corresponding to the data burst L1. Similarly, the shaped data burst d2 has its transmission beginning at the arrival of the token T2 which occurs at a time $\delta 2$ after the start of the control packet corresponding to the data burst L2. Shaped data burst d3, on the other hand, has its transmission beginning at the arrival of token T4 which occurs at a time $\delta 3$ from the start of the control packet corresponding to the data burst L3. This is due to the fact that the token T3 arrives during the pendency of the shaped data burst d2. Since the shaped data burst d3 can not commence prior to the completion of the shaped data burst d2, token T3 is not used. Shaped data burst d4 has its transmission beginning at the arrival of token T5 which occurs at a time $\delta 4$ from the start of the control packet corresponding to the data burst L4.

The type of shaping described above regulates the average rate at which data bursts are released into the optical burst switching wavelength division multiplexing layer. This rate is equal to the reciprocal of the mean of the probability distribution F (.) used to generate the tokens. Furthermore, the randomized generation of tokens prevents synchronization among the token streams of different LSP's. This is significant in that if the token generators of two LSP's traversing the same output interface of a particular node in the optical burst switching wavelength division multiplexing network happen to be synchronized, the data bursts in these LSP's will have a high probability of colliding with each other at that interface, thereby causing excessive data losses. It is possible for the random generators for tokens at different hosts to be initialized so as to prevent inadvertent synchronization.

The proposed shaping scheme also imposes the following property of the stream of data bursts of any optical LSP.Let$\{A(t)\}_{t\geq 0}$ denote total data arriving over an optical LSP until

time t at *any node* (ingress or intermediate) in the OBS WDM network. Then,

$$A(t) - a(s) \leq A_X(t) - A_X, \text{ a.s., (almost surely), for all } t \geq s,$$

where $\{A_X(t)_{t\geq}0$ denotes total data that would have arrived on that LSP until time t at that node, if data bursts were arriving at T_0, T_1, T_2, ...

It is easy to see that the domination as in Eq. 1 holds at the output of the shaper of every optical LSP, by virtue of the shaping scheme. As data bursts traverse various nodes in the optical backbone, some of them can *only be discarded*, due to contention. Furthermore, due to inherently bufferless forwarding, the relative positions of the data bursts of any optical LSP remain unchanged even after these data bursts traverse a number of nodes. Hence, the domination as in Eq. 1 holds at the output interface of every node that a given optical LSP traverses.

A framework for resource provisioning an admission control, based on the above noted shaping scheme, is provided below and in addition, the role of this shaping scheme in improving the performance of an optical burst switching wavelength division multiplexing layer will be demonstrated below for the case of TCP/IP traffic.

The following is a description of one example of how the shaping scheme can be used for traffic engineering in an OBS WDM network. Consider an output interface (fiber) of any node in an OBS WDM network. Suppose that this node is being traversed by N data pipes (optical LSP's) with the provisional data rates of $r_1,...,r_N$, respectively. If the data bursts entering into these LSP's are shaped at the ingress using Poisson shapers (this means that the probability distribution for the time interval between the successive tokens of the ith optical LSP, denoted by $F_i(.)$, is chosen to be exponential), the following holds for their data arriving at the output interface under consideration. For all $t \geq s$,

$$Ai(t) - Ai(s) \leq AP(r_i)(t) - AP(r_i)(s),$$
$$\text{a.s, for all } 1 \leq i \leq N, \text{and} \qquad (2)$$

$$A(t) - A(s) = \sum_{i=1}^{N} A_i(t) - A_i(s)) \leq A_{P(r)}(t) - A_{P(r)}(s), \text{ a.s.} \qquad (3)$$

Here, $A_i(t)$ denotes the total data arriving over the ith optical LSP until time t, $AP(x)(t)$ denotes the data that would arrive if the data bursts were arriving according to a Poisson process of rate x, and $r = \sum_{i=1}^{N} r_i$.

Now, if $p_{actual}(r_1,\ldots,r_N)$ denotes the actual burst blocking probability at the given output interface, it is intuitively appealing to say that

$$p_{\text{actual}}(r_1,\ldots,r_N) \leq p_{\text{Poisson}}(r), \qquad (4)$$

where $p_{\text{Poisson}}(r)$ denotes the burst blocking probability at that interface if the bursts were arriving according to the Poisson process of rate r. The right hand side of Inequality 4 is given by the well known Erlang loss formula,

$$p_{\text{Poisson}}(r) - \frac{(r/\mu)^c / c!}{\sum_{i=0}^{c} r/\mu)^c / c!} \qquad (5)$$

where c is the total number of wavelengths at the output interface, and $1/\mu$ is the average packet length.

If the establishment of a new optical LSP, requiring a data rate of r_{N+1} and a burst blocking probability of r_{N+1}, is requested through a given output interface at a node in an OBS WDM network, it can be admitted if and only if

$$p_{\text{Poisson}}(r) \leq p_{N+1}, \text{ with } r = \sum_{i=1}^{N+1} r_i \qquad (6)$$

Figure 5 illustrates the burst blocking probability as a function of the offered load for in interface supporting various numbers of wavelengths. Specifically, Figure 5 shows the case for in interface supporting 16,32, 48, 64, and 80 wavelength channels. It can be seen from Figure 5, that a wavelength division multiplexing network in accordance with the present invention can be operated at high utilizations using optical burst switching when the number of wavelengths is large. For example, as illustrated in Figure 5, with a target burst blocking probability which is less than 10^{-4}, a wavelength division multiplexing network in accordance with the present invention can operate with greater than 60 percent utilizations if the number of wavelength channels is 64.

The approach to supporting IP over optical backbone consists of providing MAC layer functionalities so that the underlying optical layer (using OBS) appears to the IP layer as a reliable data pipe (see Figure 3). By virtue of the shaping scheme and the connection admission control procedure described above, the burst blocking performance of every optical LSP is guaranteed. It is also possible to render some end-to-end delay characteristics to optical

LSP as follows.

As seen from Figure 3, the various elements in the MAC layer 350 such as the burst assembler 313, the burst scheduler 324 and the shaper 335, introduce some amount of delay at the ingress of an optical LSP. The following is a description of one example of how the delay due to shaping can be controlled. Assume that the establishment of an optical LSP, requiring a data rate of r_{N+1}, a burst blocking probability of p_{N+1} and an end-to-end delay guarantee of D_{N+1}, is requested between a chosen ingress-egress pair. The actual token rate may be chosen such that:

Prob[Time interval between successive tokens >

D_1] > ϵ, and Token rate > r_{N+1}.

This value token rate is then used in Eq. 6 in place of r_{N+1} to take admission control decision.

Figure 6 shows how IP DiffServ classes can be mapped onto optical LSP's that are now reliable, as well as have certain delay characteristics.

Another quality of service dimension is to logically partition the backbone into a number of optical LSP's, each providing different levels of reliability given in terms of burst blocking probability. The reliability of different optical LSP's can then be mapped onto some cost function. This, in turn, can be used by the routing protocols to forward IP packets to appropriate edge routers depending on their quality of service needs (now in terms of loss rate). This is depicted in Figure 7.

The bottleneck output interface (fiber) of a node in an OBS WDM network supporting three OC12 wavelengths (622 Mb/s per wavelength) per output interface has been simulated. This interface is traversed by a number of optical LSP's. Ten such optical LSP's each carrying ingress-to-egress data traffic supported on TCP/IP are assumed. There are (forward paths of) 4 TCP sessions in each of these optical LSP's. The acknowledgment paths (or reverse paths) of these TCP's are taken to be lossless, and they introduce only a constant delay. Simulations were run with a simulation tool.

Each of the 40 TCP sessions is started at time instant sampled from the uniform distribution over (0 s, 1 s). Once started, all TCP sources always have data to send. For simplicity, every data burst that is assembled at the MAC layer is taken to be precisely one IP packet. The delay introduced by the reverse path of every TCP session is sampled from the uniform distribution as explained in the next section.

AS shown in Figure 5, the burst traffic offered to each LSP is shaped at the ingress. The probability distribution $F_i(.)$ used in shaping the burst traffic of the i th optical LSP is taken to be exponential with a meaning $1/r_1$. This causes the data burst arrivals at the output of

each shaper to be dominated by the Poisson process of rate r_1, and the total arrive process of data bursts at the bottleneck output interface to be dominated by the Poisson process of rate $r = \sum_{i=1}^{10} r_i$. Simple greedy and exhaustive wavelength selection policy may be used to assign the reservations to the control packets arriving at the bottleneck output interface.

Simulation experiments have been run in different regimes of the target burst blocking probability, namely, 10^{-2}, 10^{-3} and 10^{-4}. For each of these values, the total allowable load r at that output interface is calculated using Erlang loss formula (Eq. 5), with $c = 3$. r_i is then taken to be $r/10$, for $i = 1, ..., 10$. For the target burst blocking probabilities of 10^{-2}, 10^{-3} and 10^{-4}, the delay introduced by the reverse path of every TCP session is sampled from the uniform distribution over [0 ms, 1 ms), [0 ms, 25 ms) and [0 ms, 50 ms), respectively. This is done so that the actual aggregate load offered by all TCP's is not much lower than the designed throughput value. For example, 40 TCP's fail to offer an average load as large as about 311 Mb/s at the packet loss probability of 10^{-2} in the end-to-end path, if their round trip times are larger than about 1 ms. It is as fundamental fact that the TCP throughput significantly deteriorates if the end-to-end packet loss probability is much larger than the inverse square of the product of the bottleneck bandwidth and the round trip delay. And, it is clearly trivial to establish Inequality 4 if the offered average load itself is much lower than the designed value for the load. The results are shown in Table I below.

Simulation experiment	Throughput (MB/s)			Burst blocking probability		
	With shaping		Without shaping	With shaping		Without shaping
	Observed	Designed		Observed	Designed	
1	302.29	311.00	0	0.73×10^{-2}	1.72×10^{-2}	1.0
2	146.68	155.0	0	1.17×1010^{-3}	2.00×10^{-3}	1.0
3	62.20	62.20	0	0.7×10^{-5}	1.51×10^{-4}	1.0

This concludes the description of the example embodiments. Although the present invention has been described with reference to a number of illustrative embodiments thereof, it should be understood that numerous other modifications and embodiments can be devised by those skilled in the art that will fall within the spirit and scope of the principles of this invention. More particularly, reasonable variations and modifications are possible in the

component parts and/or arrangements of the subject combination arrangement within the scope of the foregoing disclosure, the drawings, and the appended claims, without departing from the spirit of the invention. In addition to variations and modifications in the component parts and/or arrangements, alternative uses will also be apparent to those skilled in the art.

What is claimed is:

CLAIMS

We claim:

1. A method of selecting the offset between data bursts and their respective control packets in an optical burst switching arrangement, the method comprising:

randomly generating a plurality of tokens;

receiving a plurality of sequentially generated data bursts;

receiving a plurality of control packets, each control packet corresponding to a respective one of said plurality of data bursts;

first outputting one of said plurality of control packets and then outputting its corresponding respective data burst at a time corresponding to the generation of the first of said plurality of tokens which occurs at a time in which no data burst is being outputted, the outputting of the data burst being offset from the outputting of its corresponding respective control packet by a time period delta.

2. The method of claim 1, wherein an average rate at which said plurality of data bursts are outputted is equal to the reciprocal of the mean of the probability distribution used to generate said plurality of tokens.

3. An apparatus for selecting the offset between data bursts and their respective control packets in an optical burst switching arrangement, the apparatus comprising:

a token generator for randomly generating a plurality of tokens;

a data burst receiver for receiving a plurality of sequentially generated data bursts;

a control packet receiver for receiving a plurality of control packets, each control packet corresponding to a respective one of said plurality of data bursts;

a transmitter for first outputting one of said plurality of control packets received by said control packet receiver and for then outputting its corresponding respective data burst received by said data burst receiver at a time corresponding to the generation of the first of said plurality of tokens by said token generator which occurs at a time in which no data burst is being outputted by said transmitter, the outputting of said data burst being offset from the outputting of its corresponding respective control packet by a time period delta.

4. The apparatus of claim 3, wherein said transmitter outputs said plurality of data bursts at an average rate which is equal to the reciprocal of the mean of the probability distribution used to generate said plurality of tokens.

5. The method of claim 1, wherein if T0= 0, T1, T2, ... denote the times of occurrences

of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for i>1) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-1)th data burst is outputted at
T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)≥L(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(ω) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

6. The method of claim 2, wherein if T0= 0, T1, T2, ... denote the times of occurrences of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for i≥ 1) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-1)th data burst is outputted at
T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)>L(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(ω) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

7. The apparatus of claim 3, wherein if T0= 0, T1, T2, ... denote the times of occurrences of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for i≥ 1) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-1)th data burst is outputted at T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)≥L(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(ω) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

8. The apparatus of claim 4, wherein if T0= 0, T1, T2, ... denote the times of occurrences of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for $i \geq 1$) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-1)th data burst is outputted at T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)≥(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(ω) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

9. The method of claim 1,wherein the plurality of tokens are randomly generated according to a Poisson process.

10. The method of claim 2,wherein the plurality of tokens are randomly generated according to a Poisson process.

11. The apparatus of claim 3, wherein said token generator randomly generates said plurality of tokens according to a Poisson process.

12. The apparatus of claim 4,wherein said token generator randomly generates said plurality of tokens according to a Poisson process.

13. The method of claim 5,wherein the plurality of tokens are randomly generated according to a Poisson process.

14. The method of claim 6,wherein the plurality of tokens are randomly generated according to a Poisson process.

15. The apparatus of claim 7,wherein said token generator randomly generates said plurality of tokens according to a Poisson process.

16. The apparatus of claim 8, wherein said token generator randomly generates said plurality of tokens according to a Poisson process.

17. An optical burst switching apparatus comprising:

a source generator for generating a plurality of data bursts and their respective control packets, each data burst being generated at a time which is offset from that of its respective control packet;

an ingress router for receiving and outputting said plurality of data bursts and their respective control packets generated by said source generator;

an egress router for receiving and outputting said output of said ingress router;

at least one intermediate node, disposed between said ingress router and said egress router, for transmitting said plurality of data bursts and their respective control packets;

wherein said source generator comprises:

a token generator for randomly generating a plurality of tokens;

a data burst receiver for receiving a plurality of sequentially generated data bursts;

a control packet receiver for receiving a plurality of control packets, each control packet corresponding to a respective one of said plurality of data bursts;

a transmitter for first outputting one of said plurality of control packets received by said control packet receiver and for then outputting its corresponding respective data burst received by said data burst receiver at a time corresponding to the generation of the first of said plurality of tokens by said token generator which occurs at a time in which no data burst is being outputted by said transmitter, the outputting of said data burst being offset from the outputting of its corresponding respective control packet by a time period delta.

18. The apparatus of claim 17, wherein said transmitter outputs said plurality of data bursts at an average rate which is equal to the reciprocal of the mean of the probability distribution used to generate said plurality of tokens.

19. The apparatus of claim 17, wherein if T0= 0, T1, T2, ... denote the times of occurrences of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for $i \geq 1$) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-1)th data burst is outputted at

T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)≥L(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(ω) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

20. The apparatus of claim 18, wherein if T0= 0, T1, T2, ... denote the times of occurrences of points of a random point process in which the time periods between occurrences of success of points (Ti-T(i-1) for $i \geq 1$) are independently and identically distributed according to a probability distribution F(.) and if T0(ω)= 0, T1(ω), T2(ω), ... denote a particular realization of the random point process and if the ith data burst arrives at a time ai, and the (i-

1)th data burst is outputted at T(ki-1)(ω), then the ith data burst is outputted at time Tki(ω) which is the first point after T(ki-1)(ω) satisfying the following inequality:

Tki(ω)-T(ki-1)(ω)≥L(i-1) and Tki(ω)≥ai,L(i-1) being the(i-1)th inputted data burst; and wherein:

δi= Tki(w) - ai, δi being the offset between the ith data burst and its corresponding respective control packet.

FIG. 1

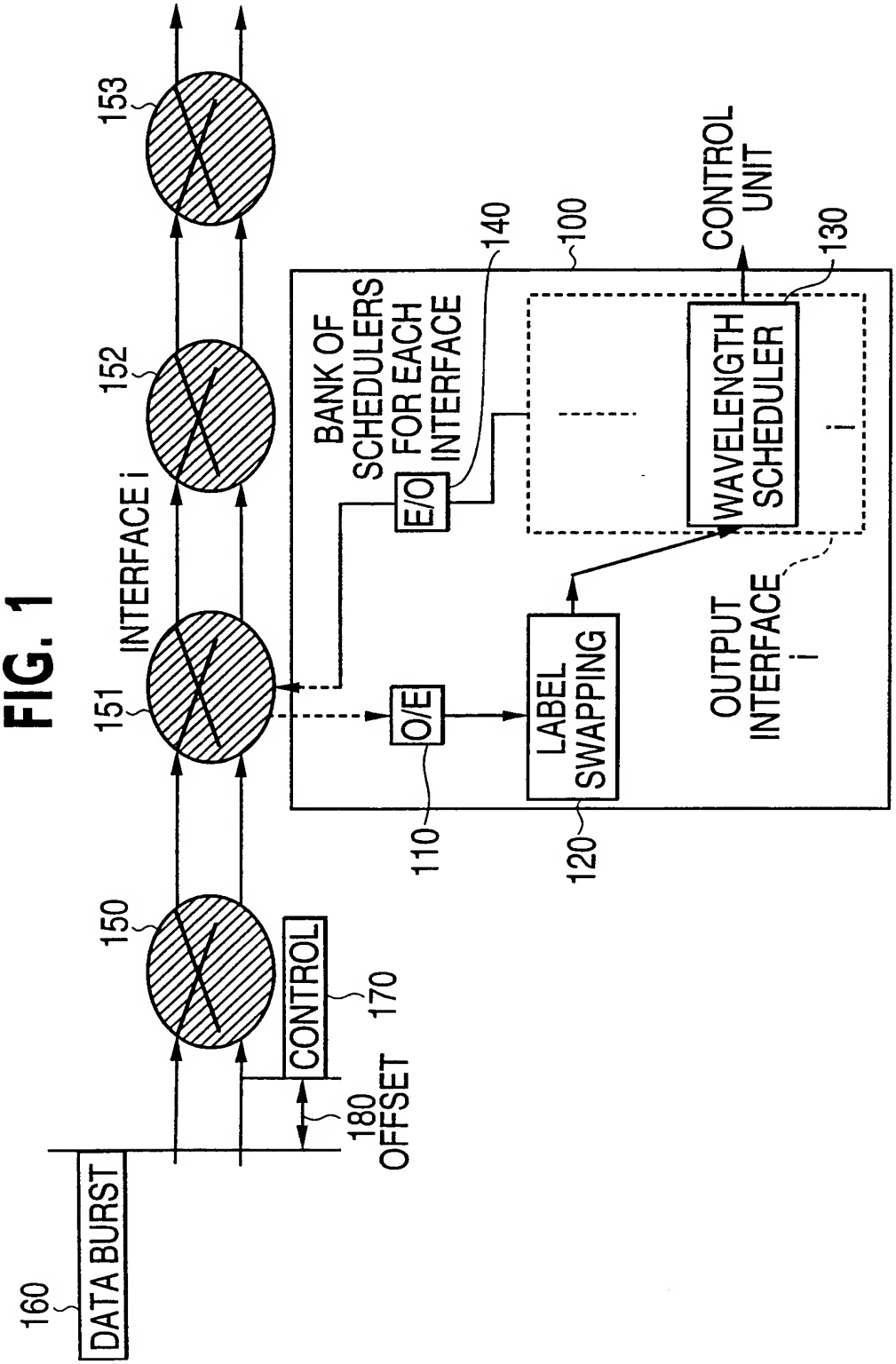

153
152
INTERFACE i
151
150
160
DATA BURST
CONTROL
170
180
OFFSET
BANK OF SCHEDULERS FOR EACH INTERFACE
E/O
140
100
CONTROL UNIT
130
WAVELENGTH SCHEDULER
i
O/E
110
LABEL SWAPPING
120
OUTPUT INTERFACE i

FIG. 2

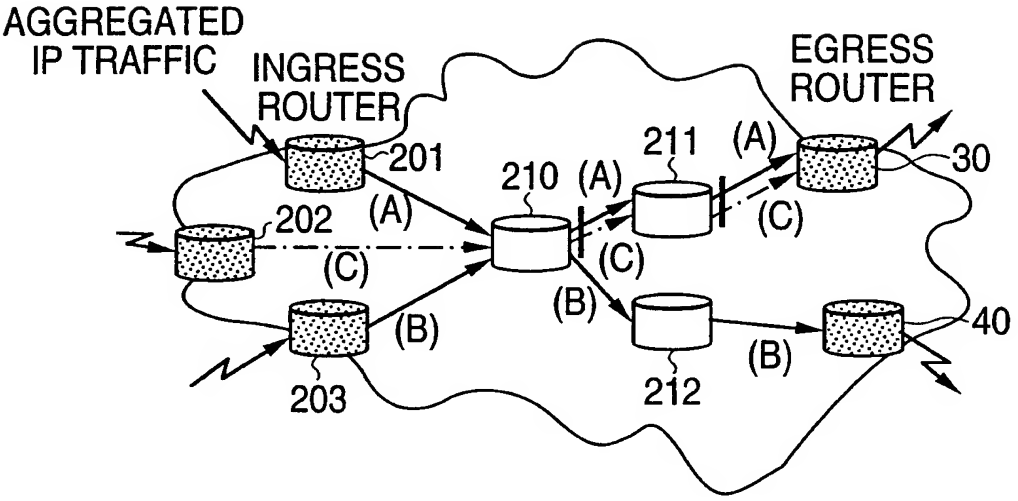

AGGREGATED
IP TRAFFIC
INGRESS
ROUTER
EGRESS
ROUTER
201
202
203
210
211
212
30
40
(A)
(B)
(C)

FIG. 3

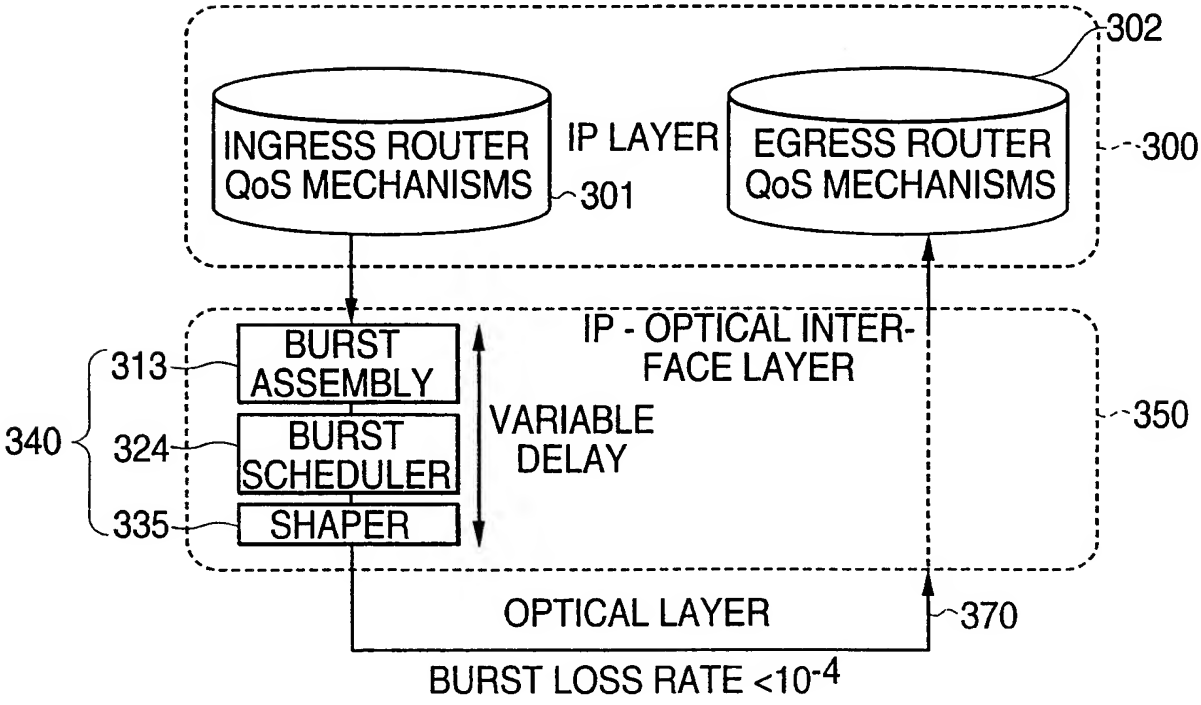

302
INGRESS ROUTER
QoS MECHANISMS
301
IP LAYER
EGRESS ROUTER
QoS MECHANISMS
300
IP - OPTICAL INTER-
FACE LAYER
350
313
340
324
335
BURST
ASSEMBLY
BURST
SCHEDULER
SHAPER
VARIABLE
DELAY
OPTICAL LAYER
370
BURST LOSS RATE <10-4

FIG. 4

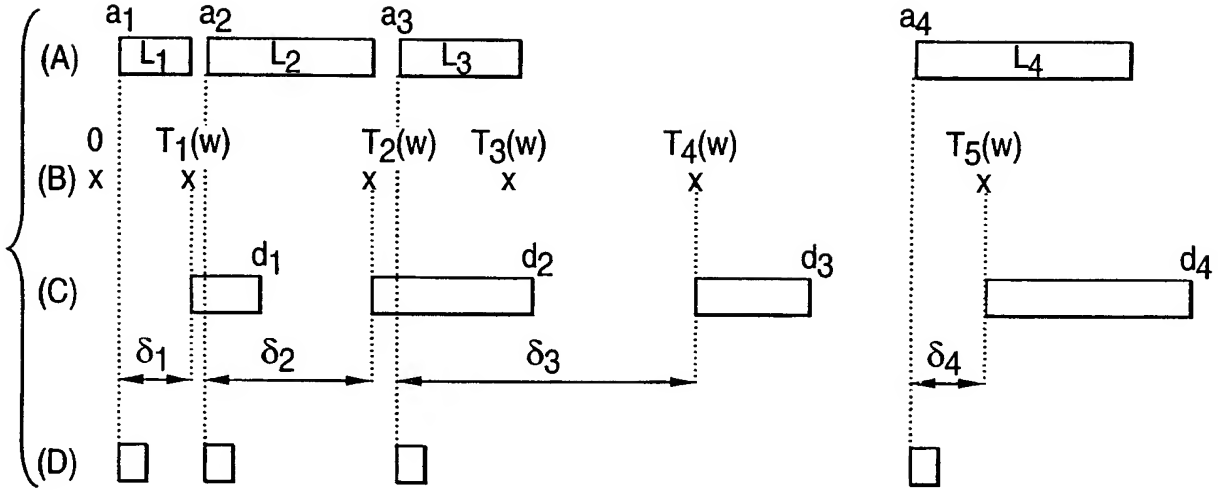

a1
a2
a3
a4
(A)
L1
L2
L3
L4
0
(B)
T1(w)
T2(w)
T3(w)
T4(w)
T5(w)
(C)
d1
d2
d3
d4
δ1
δ2
δ3
δ4
(D)

FIG. 5

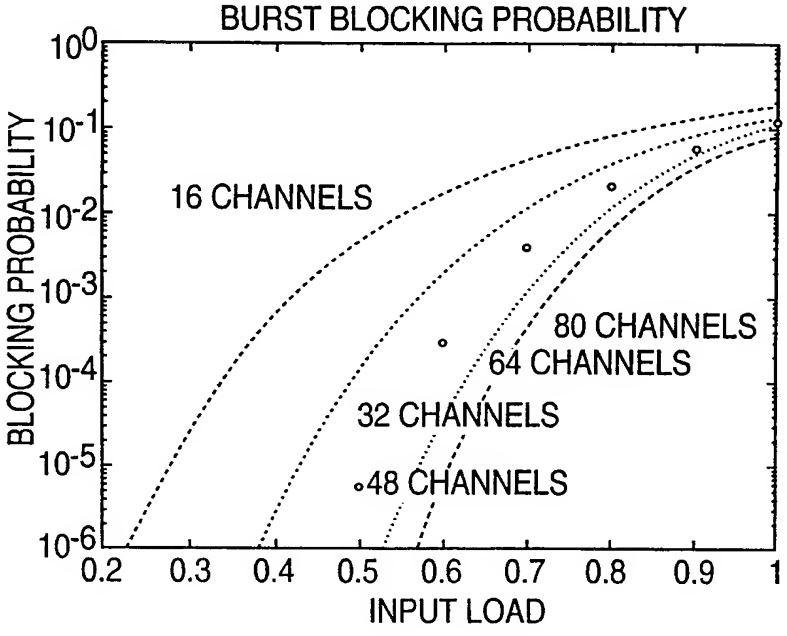

BURST BLOCKING PROBABILITY
BLOCKING PROBABILITY
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
16 CHANNELS
80 CHANNELS
64 CHANNELS
32 CHANNELS
48 CHANNELS
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
INPUT LOAD

FIG. 6

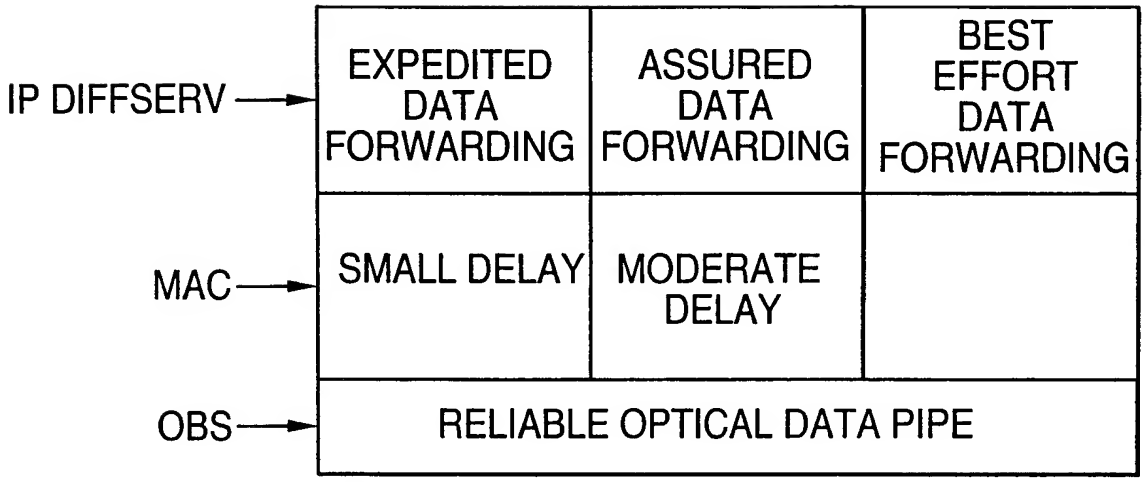
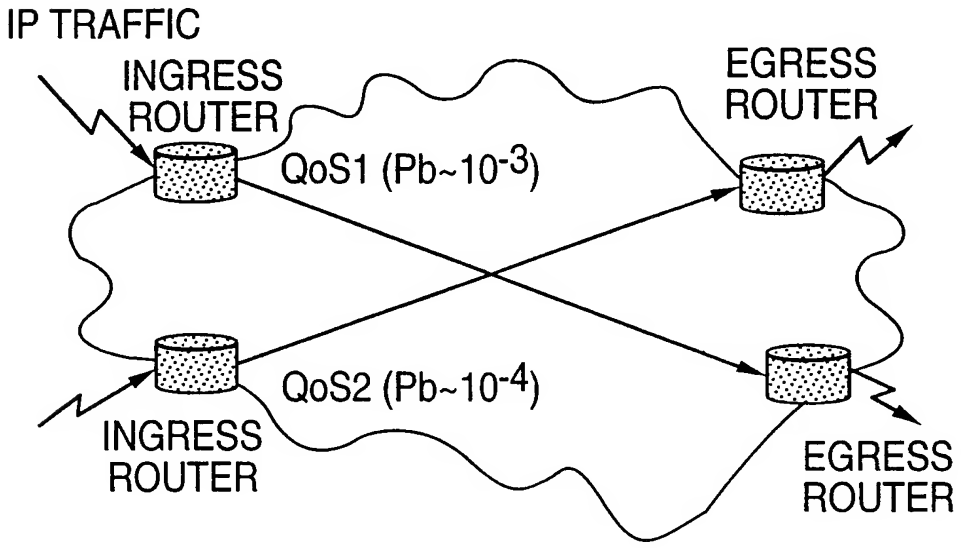

IP DIFFSERV
MAC
OBS
EXPEDITED DATA FORWARDING
ASSURED DATA FORWARDING
BEST EFFORT DATA FORWARDING
SMALL DELAY
MODERATE DELAY
RELIABLE OPTICAL DATA PIPE

FIG. 7


IP TRAFFIC
INGRESS ROUTER
EGRESS ROUTER
QoS1 (Pb~10-3)
QoS2 (Pb~10-4)
INGRESS ROUTER
EGRESS ROUTER

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 November 2001 (08.11.2001)

PCT

(10) International Publication Number
WO 01/84875 A3

(51) International Patent Classification[7]: H04Q 11/00

(21) International Application Number: PCT/US01/13508

(22) International Filing Date: 27 April 2001 (27.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/563,338 3 May 2000 (03.05.2000) US

(71) Applicant: **NOKIA, INC.** [US/US]; 6000 Connection Drive, Irving, TX 75039 (US).

(72) Inventors: **CHASKAR, Hemant**; 111 Lucust Street, Apartment 40-C-1, Woburn, MA 01801 (US). **VERMA, Sanjeev**; 2 Kimball Court, Apartment 213, Woburn, MA 01801 (US). **RAVIKANTH, R.**; 1504 Stearns Hill Road, Waltham, MA 02451 (US). **DIXIT, Sudhir, S.**; 12 Westerly Road, Weston, MA 02493 (US).

(74) Agents: **BRUNDIDGE, Carl, I.** et al.; Antonelli, Terry, Stout & Kraus, LLP, Suite 1800, 1300 N. Seventeenth Street, Arlington, VA 22209 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

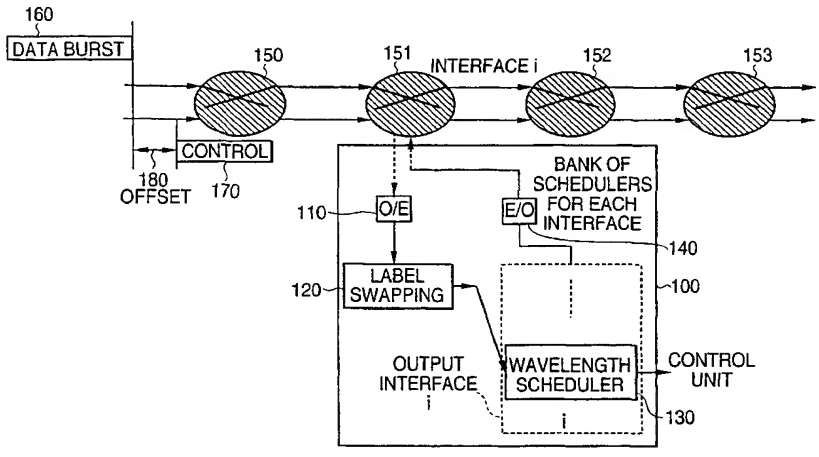
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
23 May 2002

[Continued on next page]

(54) Title: ROBUST TRANSPORT OF IP TRAFFIC OVER WDM USING OPTICAL BURST SWITCHING



(57) Abstract: A technique for selecting the offset between data bursts and their respective control packets in an optical burst switching arrangement includes: randomly generating a plurality of tokens; receiving a plurality of sequentially generated data bursts; and receiving a plurality of control packets, each control packet corresponding to a respective one of the plurality of data bursts. One of the plurality of control packets is first outputted and its corresponding respective data burst is then output at a time corresponding to the generation of the first of the plurality of tokens which occurs at a time in which no data burst is being outputted, the outputting of the data burst being offset from the output of its corresponding respective control packet by a time period. The average rate at which the plurality of data bursts are outputted may be equal to the reciprocal of the mean of the probability distribution used to generate the plurality of tokens. The plurality of tokens may be randomly generated according to a Poisson process.

WO 01/84875 A3

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 01/13508

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04Q11/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YOO M ET AL: "NEW OPTICAL BURST SWITCHING PROTOCOL FOR SUPPORTING QUALITY OF SERVICE" PROCEEDINGS OF THE SPIE, SPIE, BELLINGHAM, VA, US, vol. 3531, November 1998 (1998-11), pages 396-405, XP000950072 paragraph '0002! ---	1-20
A	QIAO C ET AL: "CHOICES, FEATURES AND ISSUES IN OPTICAL BURST SWITCHING" OPTICAL NETWORKS MAGAZINE, SPIE, BELLINGHAM, WA, US, vol. 1, no. 2, April 2000 (2000-04), pages 36-44, XP000969814 ISSN: 1388-6916 paragraph '0005! --- -/--	1-20

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 January 2002	24/01/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040. Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Meurisse, W

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 01/13508

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,A	EP 1 089 498 A (CIT ALCATEL) 4 April 2001 (2001-04-04) page 4, line 38 - line 44	1-20

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

.ormation on patent family members

Inter ›nal Application No
PCT/US 01/13508

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1089498 A	04-04-2001	AU 6131700 A	05-04-2001
		CN 1296346 A	23-05-2001
		EP 1089498 A2	04-04-2001